August 1946

BIOMETRICS BULLETIN

Vol. 2 No. 4

THE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION

STATISTICAL METHODS FOR AN INCOMPLETE EXPERIMENT ON A PERENNIAL CROP[1]

GEORGE W. SNEDECOR[2] and E. S. HABER[3]

In 1939, scarcity of labor terminated an experiment on asparagus which was planned to run for the effective lifetime of the planting—20 years or longer. Using data from the first 10 harvested crops, 1929 to 1938 inclusive, we have attempted to make some inferences about the outcome as it might have been observed in 1948 if the experiment had been continued.

Three dates of final cutting, June 1, June 15, and July 1, are the treatments on which forecasts are based. Plots receiving each treatment were laid out in six randomized blocks, the yields from which are recorded in Table 1. Progress reports on this experiment have been made from time to time (see references 4, 5, 6, and 8 at the end of this article). Moreover, some horticultural features of these forecasts have been discussed (7). The purpose of this paper is to explain the statistical methods used.

For reference, the formal analysis of variance of the 180 plot yields is given in Table 2.

Our first objective will be to examine the trends in the pair of orthogonal comparisons,

$$t_1 = b - a$$
$$t_2 = 2c - (b + a),$$

to learn if they contain information about what the future of the experiment might have shown. The quantities t_1, the differences between yields to June 15 and to June 1, are entered in the sixth column of the table; these are the yields during the first half of June. The analysis of their variance is given in the first part of Table 3. The only peculiarity of the analysis is this: since the data are differences, the factor $1^2+1^2=2$ is part of the denominator of each sum of squares.

The first question is, do the trends in these differences contain information about the future yield difference between the two cutting dates, June 15 and June 1? For each block, linear and quadratic components were calculated by the use of the ξ' orthogonal coefficients (1, 3). As examples, in block I,

$$L: -9(100) - 7(66) - 5(219) - 3(266) - (245) + (256) + 3(136) + 5(212) + 9(222) = 1{,}426$$

$$Q: 6(100) + 2(66) - (219) - 3(266) - 4(245) - 4(256) - 3(136) - (212) \quad + 2(172) + 6(622) = -1{,}233$$

The sum of the six linear components,

$$1{,}426 + 4{,}262 + \ldots + 1{,}126 = 14{,}895,$$

may be verified by applying the same coefficients to the 10 annual differences, b—a, shown in Table 1.

[1] Journal Paper No. J-1375 of the Iowa Agricultural Experiment Station, Ames, Iowa Project 890

[2] Director, Statistical Laboratory, Iowa State College.

[3] Research Professor, Horticultural Section, Iowa Agricultural Experiment Station

Examination of the annual differences leads to the conclusion that, after the first three or four years, during which the plants were reaching full productivity, no more than seasonal fluctuations occurred. Apparently the extra cuttings in the first half of June did not damage the plots during these early years: in fact, the difference in 1938 was the greatest that was observed. Clearly, there is no evidence here that cutting through June 15 was detrimental, nor is there sufficient basis for predicting future change in the differences.

For the sake of completing the analysis of variance, we calculate a sum of squares attributable to the linear trend of the annual differences during the whole 10 years (9):

$$T_1Y_L=\frac{(14{,}895)^2}{2(6)(330)}=56{,}026$$

The factor 2 in the denominator converts the sum of squares calculated from differences to the individual plot basis; 6 is due to the six replications; while 330 is the sum of the squares of the coefficients,

$$(-9)^2+(-7)^2+\ \ldots\ +(9)^2=330$$

There is often interest in testing the significance of the linear component of trend. The appropriate estimate of error is calculated from the six block differences:

$$T_1Y_LB=\frac{(1{,}426)^2+(4{,}262)^2+\ \ldots\ +(1{,}126)^2}{2(330)}-T_1Y_L$$
$$=10{,}535$$

Here, individual block data are used, no factor 6 being required in the denominator. Note the correction term, T_1Y_L. The mean square for error is 10,535/5=2,107.

To show that the sums of squares just calculated are interactions, Table 4 is constructed. The first part contains the linear regressions in each plot: thus, for the first June cutting in block 1 (see Table 1),

$$-9(201)-7(230)-\ldots+9(527)=7{,}406$$

Any of the usual schemes of computation in this table will give the interaction, $T_1Y_LB=10{,}535$: the short method used above applies to tables with only two rows (or columns). It is to be observed that the differences in the last line of the table are the linear components of $t_1=b-a$, recorded in Table 1.

That both sums of squares are interactions with Y_L is indicated in the second part of the table where each linear component is subdivided into the two sums whose difference constitutes an individual comparison with a single degree of freedom. Among the various devices for calculating interaction in this table, an easy method is to evaluate the difference between the differences,

$$(111{,}421-57{,}147)-(81{,}487-42{,}108)=14{,}895$$

This makes it clear that the quantity calculated above, $\frac{(14{,}895)^2}{2(6)\ (330)}=56{,}026$, is the interaction between the treatments, a and b, and the two parts of the linear regressions, those above the origins and those below.

For testing significance of the linear trend, F=56,026/2,107=26.59, which for degrees of freedom 1 and 5 is highly significant. But in the asparagus data, it is believed that this trend did not persist after the first four years, the trend during the last six years being non-significant. Whence the conclusion reached before, that the linear trend is not a secure basis for forecasting.

Deviations from the linear trend may be fitted by a second degree (parabolic) curve, the results being the quadratic components calculated earlier and entered in Table 1. Except in block III, the value of Q is negative, indicating a tendency for t_1 to decrease in the latter part of the 10-year period (the 1938 difference is usually an exception). The test of significance may be had by calculating:

$$T_1Y_Q=\frac{(-2{,}554)^2}{2(6)\ (330)}=4{,}118$$

$$T_1Y_QB=\frac{(-1{,}233)^2+\ \ldots\ +(-375)^2}{2(132)}-T_1Y_Q$$
$$=11{,}602$$

Two of the divisors are the same as before, but the sum of the squares of the coefficients is now

$$(6)^2+(2)^2+(-1)^2+\ \ldots\ +(6)^2=132$$

Since the mean square for error is 11,602/5=2,320, F=4,118/2,320=1.77, d.f.=1 and 5, non-significant.

If this quadratic trend were a population characteristic, it would show that the yields during the first half of June became smaller and smaller until the yield from the plots cut until June 1 surpasses that from the June 15 plots. This would denote injury due to the

longer cutting period. Actually, the evidence is not convincing—what would have been the future of t_1 is not clear from the first 10 years' experience.

With the data now available the analysis of variance of t_1 is completed in the last part of Table 3. The remaining components in the trend are significant, F, being 4,626/579=7.99. There is no occasion to isolate them because the variation evidently is due to differences in yearly yields.

In contrast to the foregoing, $t_2=2c-(a+b)$, which is twice the difference between the yield to July 1 and the average to the two earlier dates, has undoubted trends in the population. Table 1 shows that these differences changed from positive to negative when the experiment was about half completed. The computations with t_2 follow the same pattern as those for t_1 with a single distinction: the first divisor, 2, is replaced throughout by $2^2+1^2+1^2=6$. The analysis of variance for t_2 is shown in Table 5. There is no question that both the linear and quadratic trends are characteristic of the populations sampled. The latter trend shows that the differences were becoming wider each year. It is clear that cutting to July 1 depletes the vitality of the plants to such an extent that little further yield can be expected from these plots.

It should be observed that the totals for years in Table 3 and 5 add to the interaction of treatment by years in Table 2:

92,528+624,815=717,343

Also, that the totals for error have the sum, BTY:

42,416+62,511=104,927

One criterion for deciding upon the date to which cutting may be continued is the total yield of the planting throughout its life. In this experiment, the date for maximum yield is evidently near June 15. To determine this optimum date more precisely, the annual treatment yields are accumulated in Table 6; the yield in 1931, for example, being the sum of the yields for that year and the two preceding. In the last line are the total yields for 10 years.

These cumulative yields show the superiority of the June 15 cutting date. In 1936 the total of the crops cut to July 1 fell below that to June 15: the heavier cutting had killed some of the plants and had impaired the vigor of others.

In each year a parabola was fitted to the cumulative totals for the three cutting treatments. Using the summation method (2) (8), the equation is

$$Y=\frac{a+b+c}{3}+\frac{c-a}{2}X+\frac{a-2b+c}{2}(X^2-\tfrac{2}{3}),$$

where a, b and c are the cumulative totals corresponding to the dates June 1, June 15 and July 1; and X (assumed equally spaced) has the three values, —1, 0, +1, for these successive dates. The maximum yield is

$$Y(\max)=b+\frac{(c-a)^2}{8(2b-a-c)}$$

at the date,

$$X(\max)=\frac{c-a}{2(2b-a-c)}$$

During the early years of the experiment the parabolas had no maxima because the prolonged cutting had not sufficiently decreased yield in the July 1 plots. Beginning in 1932, the optimum dates and maximum yields, computed from the formulas above, are recorded in Table 7. The optimum dates can be made the basis for some useful forecasts.

If the optimum dates are plotted against years the resulting curve will be found hyperbolic in form. This may be verified by plotting the reciprocals of the dates as ordinates with years as abscissas, the equation of the resulting straight line being calculated as

$$\frac{1}{D}=1.934+0.6125Y,$$

where D is optimum date and Y is the year expressed as deviation from 1935. Using this as a forecasting equation, optimum dates are estimated for years succeeding the termination of the experiment. Three of these follow:

1940, June 19
1943, June 18
1955, June 17

For the twentieth harvest (1948), the optimum date for cessation of cutting is estimated as June 17.5 with the 95% confidence interval, June 16.5 to June 20.

It is plain that June 17 is close to the optimum date for plantings to be harvested for 20

years or more. In reference (7) it is shown that this seems to be independent of soil fertility so far as the experiment at Ames supplies competent evidence. If these findings prove generally true, managers of canning factories may plan for simultaneous termination of the asparagus pack from all plantings at a date to be determined for each climatic region.

TABLE 1A

Yield of asparagus (ounces) from 18 plots for 10 years. Three dates of final cutting.

		6/1	6/15	7/1		
Block	*Year*	*a*	*b*	*c*	*b—a*	*2c—(b+a)*
I	29	201	301	362	100	222
	30	230	296	353	66	180
	31	324	543	594	219	321
	32	512	778	755	266	220
	33	399	644	580	245	117
	34	891	1147	961	256	—116
	35	449	585	535	136	36
	36	595	807	548	212	—306
	37	632	804	565	172	—306
	38	527	749	353	222	—570
Total		4760	6654	5606	1894	—202
				L:	1426	—14450
				Q:	—1233	—3127
II	29	185	236	341	51	261
	30	216	256	328	40	184
	31	317	397	487	80	260
	32	448	639	622	191	157
	33	361	483	445	122	46
	34	783	998	802	215	—177
	35	409	525	478	116	22
	36	566	843	510	277	—389
	37	629	841	576	212	—318
	38	527	823	299	296	—752
Total		4441	6041	4888	1600	—706
				L:	4262	—16504
				Q:	—40	—3098
III	29	209	226	357	17	279
	30	219	212	354	—7	277
	31	357	358	560	1	405
	32	496	545	685	49	329
	33	344	415	520	71	281
	34	841	833	871	—8	68
	35	418	451	538	33	207
	36	622	719	578	97	—185
	37	636	735	634	99	—103
	38	530	731	413	201	—435
Total		4672	5225	5510	553	1123
				L:	2751	—12615
				Q:	+896	—3812

		6/1	6/15	7/1		
Block	*Year*	*a*	*b*	*c*	*b—a*	*2c—(b+a)*
IV	29	219	330	427	111	305
	30	222	301	391	79	259
	31	348	521	599	173	329
	32	487	742	802	255	375
	33	372	534	573	162	240
	34	773	1051	880	278	—64
	35	382	570	540	188	128
	36	505	737	577	232	—88
	37	534	791	524	257	—277
	38	434	614	343	180	—362
		4276	6191	5656	1915	845
				L:	2077	—12885
				Q:	—1076	—2832
V	29	225	307	382	82	232
	30	239	297	321	58	106
	31	347	463	502	116	194
	32	512	711	684	199	145
	33	405	577	467	172	—48
	34	786	1066	763	280	—326
	35	415	610	468	195	—89
	36	549	779	548	230	—232
	37	559	741	621	182	—58
	38	433	706	352	273	—435
		4470	6257	5108	1787	—511
				L:	3253	—10261
				Q:	—726	+244
VI	29	219	342	464	123	367
	30	216	287	364	71	225
	31	356	557	584	201	255
	32	508	768	819	260	362
	33	377	529	612	152	318
	34	780	969	1028	189	307
	35	407	526	651	119	369
	36	595	772	660	177	—47
	37	626	826	673	200	—106
	38	518	722	424	204	—392
		4602	6298	6279	1696	1658
				L:	1126	—10648
				Q:	—375	—4813

Entered as second-class matter, May 25, 1945, at the post office at Washington, D. C., under the Act of March 3, 1879. The Biometrics Bulletin is published six times a year—in February, April, June, August, October and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N.W., Washington 6, D. C.

Membership dues in the American Statistical Association are $5.00 a year, of which $3.00 is for a year's subscription to the Quarterly Journal, fifty cents is for a year's subscription to the ASA Bulletin and members who pay $1.00 additional receive a year's subscription to the Biometrics Bulletin. Dues for Associate members of the Biometrics Section are $2.00 a year, of which $1.00 is for a year's subscription to the Biometrics Bulletin. Single copies of the Biometrics Bulletin are 60 cents each and annual subscriptions are $2.00 Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N.W., Washington 6, D. C.

TABLE 1B

Annual yields of asparagus (ounces) from all plots.

Year	a	b	c	b—a	2c—(b+a)
29	1,258	1,742	2,333	484	1,666
30	1,342	1,649	2,111	307	1,231
31	2,049	2,839	3,326	790	1,764
32	2,963	4,183	4,367	1,220	1,588
33	2,258	3,182	3,197	924	954
34	4,854	6,064	5,305	1,210	—308
35	2,480	3,267	3,210	787	673
36	3,432	4,657	3,421	1,225	—1,247
37	3,616	4,738	3,593	1,122	—1,168
38	2,969	4,345	2,184	1,376	—2,946
	27,221	36,666	33,047	9,445	2,207

For t_1; L=14,895, Q=—2,554. For t_2; L=—77,363, Q=—17,438.

TABLE 2

Analysis of variance of yields of asparagus

Source of Variation	Degrees of Freedom	Sum of Squares
Blocks (B)	5	102,532
Treatments (T)	2	756,930
BT	10	144,647
Years (Y)	9	5,461,360
BY	45	122,888
TY	18	717,343
BTY	90	104,927
Total		7,410,627

TABLE 3

Analysis of variance of the yields during the first half of June, t_1=b—a.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Blocks	5	66,103	
Years	9	92,528	
Error	45	42,416	
Completed Analysis			
Years:			
Linear, T_1Y_L	1	56,026	56,026
Quadratic, T_1Y_Q	1	4,118	4,118
Remainder	7	32,384	4,626
	9	92,528	
Error:			
Linear, T_1Y_LB	5	10,535	2,107
Quadratic, T_1Y_QB	5	11,602	2,320
Remainder	35	20,279	579
Total	45	42,416	

TABLE 4

A. Linear components of trends in yield of June 1 and June 15 cuttings on six blocks of asparagus experiment.

Cutting Date	1	2	3	4	5	6	Total
June 1	7,406	7,519	7,396	4,990	5,212	6,856	39,379
June 15	8,832	11,781	10,147	7,067	8,465	7,982	54,274
Differences	1,426	4,262	2,751	2,077	3,253	1,126	14,895

B. The positive and negative parts of the linear comparison in each cutting treatment, June 1 and June 15. (From annual yields in table 1)

Parts of Regression Comparison (illustrated by June 1 data)	June 1	June 15
9(2,969) +7(3,616) + . . . + (4,854)	81,487	111,421
(2,258) +3(2,963) + . . . +9(1,258)	42,108	57,147

TABLE 5

Analysis of variance of the differences, $t_2 = 2c - (a+b)$.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Blocks	5	78,545	
Years:			
T_2Y_L	1	503,791	503,791
T_2Y_Q	1	63,991	63,991
T_2Y_D	7	57,033	8,148
Total	9	624,815	
Error:			
T_2Y_LB	5	13,893	2,779
T_2Y_QB	5	18,271	3,654
T_2Y_DB	35	30,347	867
Total	45	62,511	

TABLE 6

Cumulative yields (ounces per plot) for three cutting treatments in each of 10 years.

Year	June 1	June 15	July 1
1929	210	290	389
1930	433	565	741
1931	775	1,038	1,295
1932	1,269	1,736	2,023
1933	1,645	2,266	2,556
1934	2,454	3,276	3,440
1935	2,867	3,821	3,975
1936	3,439	4,597	4,545
1937	4,042	5,387	5,144
1938	4,537	6,111	5,508

TABLE 7

Optimum dates and maximum cumulative yields calculated from fitted parabolas

Year	Optimum Date	Cumulative Yield
	X(max)	Y(max)
1932	2.100 (7/18)	2,132
1933	1.376 (7/7)	2,579
1934	0.748 (6/27)	3,461
1935	0.692 (6/26)	4,013
1936	0.457 (6/23)	4,724
1937	0.347 (6/21)	5,482
1938	0.223 (6/19)	6,165

REFERENCES

1. Anderson, R. L. and Houseman, E. E. Tables of Orthogonal Polynomial Values Extended to N equals 104. Iowa Agricultural Experiment Station Research Bulletin 297, Iowa Agricultural Experiment Station (1942).
2. Fisher, R. A. Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh (1923-1944).
3. Fisher, R. A. and Yates, F. Statistical Tables. Oliver and Boyd, Edinburgh 1938, 1942).
4. Haber, E. S. Effect of Size of Crown and Length of Cutting Season on Yields of Asparagus. Journal of Agricultural Research, 45:101-109 (1932).
5. Haber, E. S. Effect of Harvesting, Spacing and Age of Plants on Yields of Asparagus. Iowa Agricultural Experiment Station Bulletin 339. (1935).
6. Haber, E. S. Effect of Spacing and Length of Harvesting on Yields of Asparagus. Iowa Agricultural Experiment Station Bulletin, P. 6 (1940).
7. Haber, E. S. and Snedecor, George W. Forecasts from an Incomplete Experiment on Asparagus. Submitted for publication in the American Society for Horticultural Science Proceedings (1946).
8. Snedecor, George W. Statistical Methods, 4th Edition, pages 414-421. Collegiate Press, Inc., Ames Iowa (1946).
9. Yates, F. and Cochran, W. G. The Analysis of Groups of Experiments. Journal of Agricultural Science, 28:556-580 (1938).

A CRITICAL STUDY OF THE SUMMATION-OF-DIFFERENCE-IN-RANK METHOD OF DETERMINING PROFICIENCY IN JUDGING DAIRY PRODUCTS

W. D. Baten and G. M. Trout**

Michigan Agricultural Experiment Station

Prior to 1937 the summation of differences in rank between the official and student placements of samples of butter, cheese, milk or ice cream was used, in part, to determine the rating of the student judges in the collegiate students national dairy products judging contests. Since then this method has not been used for that purpose, as it was pointed out that equal sums of differences in placements of samples might actually represent a wide range of coefficients of correlation in placement (1). Despite this observation and change of evaluating proficiency in dairy products judging, the method seems to have merit in the minds of some as a relatively quick means of segregating good judges from poor judges, and also to determine how close the student ranking of a series of samples is to the official ranking. In view of this a further study of the method seemed desirable. Consequently, a study was made to ascertain at what level the summation of differences in ranks represented judging ability or mere guessing.

* Journal article No. 802(N.S.) Michigan Agricultural Experiment Station, East Lansing, Mich.

** Professor of Mathematics and Professor of Dairy, Michigan State College.

Scores obtained from mathematical possibilities of placements.

The question naturally arises as to whether or not the individual judge is using his knowledge of the dairy products or is making decisions by mere guessing or by chance. To be able to answer this question, it is necessary to know the score which can be obtained on the average by chance alone, or by people who know nothing about the materials they are assorting.

The score on placement is arrived at as follows: In judging dairy products each student judge is required to examine the materials and place a score on them and frequently is asked to assort them in ascending order from the best to the poorest. Experienced judges have agreed upon the ranks before hand. By the method under study, the score an individual judge receives is obtained by adding the differences between the official ranks and the ranks he gives to the materials. For example, in the placing of seven samples, if the

official ranks are	1, 2, 3, 4, 5, 6, 7
and the student's ranks are	3, 2, 1, 6, 4, 5, 7
the differences are	2, 0, 2, 2, 1, 1, 0

then his score is

$$2+0+2+2+1+1+0=8$$

Hence, the individual with the lowest score, or the team with the lowest score, is the winner.

There are 5,040 ways of assorting or arranging 7 different items. The scores which may be obtained from these ways of ranking the products run from 0 to 24. Odd scores are not possible. If a person gets all of the ranks, as the official ranks, his score is 0; if he gets none of his ranks correct, his score may be as high as 24. There are 251 ways of getting a score of 24, 361 ways of getting a score of 22, and so on as shown in Table 1.

It is interesting to note that more than 60 per cent of the scores are 16 and greater; that more than one-fourth of them are equal to 20, 22 and 24; and that less than 6 per cent are 8 or less. Less than one-fourth are equal to or less than 12. This distribution as plotted in Fig. 1 enables one to see where the majority of the scores are located with reference to the perfect or zero score. This distribution of scores is skew to the left. The mean, standard deviation, and skewness of these scores which

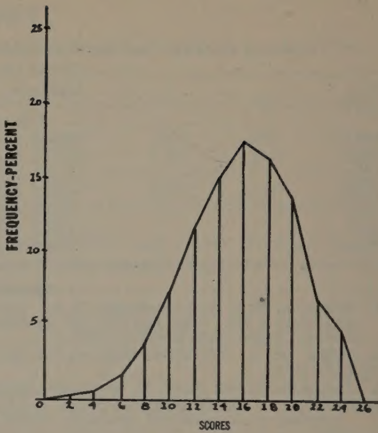

Fig. 1.

Showing the distribution of possible sums of differences in ranks between official and calculated placements of seven samples.

Table 1. *Distribution of all possible scores made from the 5040 different rankings of seven samples from 1 to 7.*

Scores	*Frequency*	*Frequency as Percentage*	*Cumulative Frequency as Percentage*
0	1	.02	.02
2	6	.12	.14
4	25	.50	.64
6	76	1.51	2.15
8	187	3.71	5.86
10	358	7.10	12.96
12	593	11.76	24.72
14	760	15.08	39.80
16	884	17.54	57.34
18	830	16.47	73.81
20	708	14.05	87.86
22	361	7.16	95.02
24	251	4.98	100.00
	5040	100.00	

might be obtained by chance are respectively.

Mean	15.99
Standard deviation	4.31
Skewness	0.24

The mean or arithmetic average of 15.99 or 16 shows that if a person who did not know any-

thing about the dairy product which he was judging, assorted seven samples by chance alone, his score would be on the average about 16; that about 75 per cent of these chance scores would be between 12 to 20 inclusive and that about 44 per cent of the scores would be between 12 and 16 inclusive.

If significance is placed at the level of 1 in 20, then a person assorting these materials should make a score of 8 or less before it could be said that he was not guessing. A score of 12, 14, 16 or 18 is what could be expected, on the average, of people who guessed at the ranks. Thus a three-person team score of 30 does not reveal any *extra-ordinary* ability at judging a series of seven samples.

Correlation between official and student ranks

Rank correlation coefficients between the official assortings and the various assortings were calculated. These coefficients might have values from —1 to +1. If the student's ranks were the same as the official ranks, the correlation coefficient would be 1, and if they were in reverse order, the coefficient would be —1. There are many assortments which lead to correlation coefficients which are near zero. The distribution of these rank correlation coefficients is given in Table 3.

Table 3. *Distribution of the rank correlation coeficients between the official ranks and the possible ranks.*

Classes of correlation coefficients	*Distribution of correlation coefficients*	
	Number	*Percent*
—1.000 to —0.875	86	1.71
—0.875 to —0.625	266	5.28
—0.625 to —0.375	644	12.78
—0.375 to —0.125	976	19.36
—0.125 to +0.125	1096	21.75
0.125 to 0.375	985	19.54
0.375 to 0.625	638	12.66
0.625 to 0.875	321	6.36
0.875 to 1.000	28	.56
	5040	100.00

Mean = —0.003
Standard Deviation = 0.410

This distribution is shown graphically in Fig. 2. The graph is nearly symmetrical. This distribution is not normally distributed. Most of the items are between —0.375 and 0.375. A correlation coefficient is significant if it is either greater than 0.75 or less than —0.75, meaning that most of the assortments which can be made do not indicate very much knowledge of the dairy products which are being judged. Less than 7 per cent of these rank correlation coefficients are less than or equal to —0.750 and greater than or equal to +0.750. This distribution of correlation coefficients shows that a large majority of these coefficients are not significantly greater than zero.

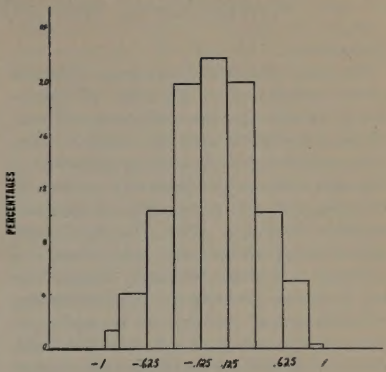

Fig. 2

Distribution of all possible rank correlation coefficients in assorting seven samples.

Summary

Grading an individual in dairy product judging proficiency by the summation-of-difference-in-rank method has little value unless the individual, in placing seven samples, can score 8 or less. The reason this method is of little value is because there are so many ways of making large scores as compared to those for making small scores. A person who knows nothing about judging the product can on the average make a score which appears to show judging ability which does not exist. Over 93 per cent of the rank correlation coefficients are not significant.

LITERATURE CITED

1. Trout, G. M., and Sharp, P. F. 1937. The reliability of flavor judgements, with special reference to the oxidized flavor of milk, N. Y. (Cornell) Agr. Exp. Sta. Memoir 204 60 pp. p. 19.

STATISTICS AT THE UNIVERSITY OF WISCONSIN

Instruction

The principal aim of the program of instruction in statistics at the University of Wisconsin is to acquaint potential users and consumers of statistics with the statistical viewpoint, and to give them an appreciation of the more common problems and methods of sampling, reduction and analysis of data, and statistical inference. Thus the aim is not production of mathematical statisticians and statistical consultants, although students who are interested primarily in the development of mathematical statistics or in serving as consultants in statistical work may avail themselves of the Master's and Doctor's degrees in mathematics with emphasis on mathematical statistics, which are offered by the Department of Mathematics and presuppose an undergraduate major in mathematics. This Doctor's degree may carry a minor in applied statistics involving courses taken, under the supervision of the Division of Statistics, in several other departments.

In recognition of the fact that persons with little or no research experience more readily acquire the statistical viewpoint and an appreciation of statistical procedures when the problems discussed and the illustrative data employed are drawn from familiar fields, than when statistics is taught as a separate discipline, a student generally makes his first contact with statistics at the University of Wisconsin in an introductory course offered by his department, or by a related department. According to his needs and mathematical proficiency, this course may then be followed by intermediate and advanced courses in statistical methods offered by his own or other departments, or by courses in mathematical probability and mathematical statistics offered by the Department of Mathematics, or by a combination of these. A student whose contact with statistics promises to be closer than that of a mere consumer is urged to include a year course in calculus in his program at an early stage, since knowledge of differential and integral calculus is a prerequisite for the courses in mathematical probability and mathematical statistics offered by the Department of Mathematics, and is a definite asset in the advanced courses in statistics offered by the other departments.

As a result of the greatly expanded use of statistical methods within the past few years not only in the social sciences but also in the biological and physical sciences, and in engineering and industry as well, more departments of instruction have introduced elementary courses in statistics, while those with such courses have added work in statistics at the intermediate, advanced, and postgraduate levels. In several departments one or more statistics course(s) is (are) required for the undergraduate degree, and practically all postgraduate study in the social sciences at the University of Wisconsin involves statistical methods in one form or another.

Recent developments at the University of Wisconsin include additional and expanded statistics courses in commerce, economics, education, geography, mathematics, medicine, psychology, sociology, and social work. More specifically statistical instruction has been expanded in such areas as: actuarial science, commercial research, industrial psychology, industrial statistics, industrial quality control and acceptance inspection, personnel administration, public opinion analysis, social service, statistics in government service, test construction and others. A complete list of courses in statistics and allied topics currently offered by the University of Wisconsin is given at the end of this article.

As members of the Division of Statistics, faculty personnel in charge of three courses, whatever their departments, meet on a common ground to discuss the contents of existing and proposed courses, applications of statistical methods, and formulation of theoretical concepts. Such discussion is greatly facilitated by the fact that all Colleges and Schools of the University of Wisconsin are located on a single campus. It is recognized that, while courses in statistical technique at the introductory level may vary in the emphasis placed on particular techniques and in the nature of the illustrative materials, an advanced study of the subject must be based on a reasonable proficiency in mathematics. As a consequence instruction and research

in the Department of Mathematics has been enlarged in the area of mathematical probability and theoretical statistics, and simultaneously other departments have drawn on the mathematician to supplement and enrich their work, especially at the advanced and postgraduate levels.

Division of Statistics

The Division of Statistics of the University of Wisconsin is an interdepartmental organization created in 1941 for the purpose of (1) coordinating statistics courses throughout the University, (2) arranging a broader offering in statistics for graduate students who plan to enter teaching or research, (3) providing an opportunity for a graduate student to acquire a thorough mathematical background suitable for an appreciation of the assumptions and limitations of the statistical methods particularly suited to his major field, and (4) supervising a minor for the Degree of Doctor of Philosophy wherein the selection of courses in statistics may involve instruction in several departments. The Division of Statistics serves in an advisory capacity (a) to the Dean of the Graduate School with respect to the program of study of a candidate for the Degree of Doctor of Philosophy with a minor in statistics that involves statistics courses taken in several departments, and (b) to the department and the dean of the college or school concerned in cases of proposed new courses, or revisions of existing courses, in statistics.

Membership in the Division is on an individual basis rather than departmental. The present 23 members are distributed among the departments of the University as follows: Agricultural Bacteriology, 1; Agricultural Economics, 3; Agronomy, 1; Botany, 1; Commerce and Economics, 2; Education, 4; Genetics, 2; Geography, 1; Mathematics, 4; Physiology, 1; Psychology, 1; Sociology, 2. The officers of the Division are a chairman and a secretary, elected annually. (The officers for 1945-46 are Professor Erwin A. Gaumnitz, chairman; Associate Professor Churchill Eisenhart, secretary.)

There are two committees of the Division: the Executive Committee, one of the functions of which is to review applications for approval of proposed new courses, or revision of existing courses, and send recommendations on behalf of the Division to the dean and department concerned; the Program Committee, the principal function of which is to review, in conjunction with the major professor concerned, the program of study submitted by a candidate for the Doctor of Philosophy with a minor in statistics that involves courses taken in several departments, and to report a joint recommendation to the Dean of the Graduate School. The regulations of the Graduate School permit a diversified minor involving instruction in more than one department, with the courses to be taken by the candidate and the method of testing the candidate stipulated by a special committee of three or more members, including the major professor, appointed by the Dean of the Graduate School. The Dean of the Graduate School has expressed willingness to accept in general the joint recommendation of the Program Committee of the Division and the major professor in lieu of the recommendation of a special committee appointed by him.

For admission, under the supervision of the Division, to a diversified minor in statistics the candidate must have had at least one semester of calculus and an introductory course in statistics acceptable to the Division. The minor consists of a minimum of twelve cerdits above the introductory level, including Mathematics 137a and Mathematics 137b, or their equivalent. The courses to be taken by a particular student will depend upon his major field and his individual interests and needs.

Consultation and Service

There are three members of the faculty of the University of Wisconsin who serve, on a part-time basis, as consultants in statistics to the staff of the University: Drs. Kenneth J. Arnold, Churchill Eisenhart, and James H. Torrie. Broadly speaking, Dr. Arnold is responsible for consultation relating to the physical sciences and to engineering; Dr. Eisenhart, for statistical consultation relating to the biological sciences and to agricultural experimentation. Dr. Torrie cooperates in the statistical aspects of planning of agricultural experiments, especially field trials. At present there is no one specifically responsible for

statistical consultation relating to the social sciences, although the University intends to have such a person in the future.

In 1929 the University Computing Service was established in the Department of Mathematics for the purpose of providing computational assistance to staff of the University in long or complex calculations. Its functions now include (a) provision of counsel in the planning of lengthy or complex calculations and computational assistance in their execution, (b) provision of information regarding existing mathematical tables and other aids to computation, and (c) construction of special tables and aids to computation. The University Computing Service currently employs a single part-time computer, and operates under the supervision of Drs. Arnold and Eisenhart. It is expected that in the near future the demands on the Service will again reach prewar level, and the Service will again employ two or more assistants. Since the establishment of the Agricultural Statistical Service, the two Services have worked in close cooperation, and now the Agricultural Statistical Service assumes the responsibility for those computations arising in the Agricultural Experiment Station that would otherwise be handled by the University Computing Service.

The Agricultural Statistical Service was created in 1937 as a branch of the Office of the Associate Director of the Agricultural Experiment Station of the University of Wisconsin. Its activities include: (a) provision of counsel in the designing of experiments and the planning of inquiries; (b) provision of computational and interpretive assistance in the statistical analysis of research data from experiments and inquiries conducted by the Station, or reported in the literature and bearing on research of the Station; (c) study of the mthods of measurement, sampling, etc., employed by the research staff of the Station, with a view to recommending improvements where feasible and desirable; (d) research in statistical methodology and applications; and (e) preparation of memoranda to keep the staff of the Station informed of developements in statistical theory and methodology that bear directly on the work of the Station. The Agricultural Statistical Service operates under the direction of the Station Statistician, Dr. Eisenhart, and employs in addition one full-time and one part-time assistant.

In the fiscal year 1946-47 it is planned to combine the Agricultural Statistical Service and the Office of the Consulting Physicist into a single administrative unit, the Biometry and Physics Section of the Agricultural Experiment Station.

Personnel

Faculty personnel in charge of one or more courses in statistical technique or statistical theory include Professors D. S. Anderson (Agricultural Economics), A. S. Barr (Education), H. P. Evans (Mathematics), P. G. Fox (Commerce and Economics), E. A. Gaumnitz (Commerce and Economics), and T. C. McCormick (Sociology); Associate Professors C. Eisenhart (Mathematics), and J. H. Torrie (Agronomy); Assistant Professors K. J. Arnold (Mathematics), and D. A. Grant (Psychology).

K. J. Arnold and C. Eisenhart, together with Ruth C. Bernstein, Student Computer, form the staff of the University Computing Service. The staff of the Agricultural Statistical Service consists of C. Eisenhart, Statistician, and (Mrs.) Dorothy M. Brill and R. J. Hader, Assistants.

Courses in Statistics and Related Topics Offered by the University of Wisconsin 1945-46

INTRODUCTORY STATISTICAL COURSES

Agronomy 136.	(Math. 36). Biometry
Commerce 31.	Business Statistics
Economics 30.	Economic Statistics
Economics 130.	Statistical Methods
Education 112.	Statistical Methods Applied to Education
Mathematics 134.	Mathematics of Elementary Statistics
Psychology 130.	Psychometric Methods
Sociology 132.	Introductory Social Statistics

No more than one of the courses in this group may be taken for full undergraduate credit, and none may be counted as graduate work in statistics.

FURTHER COURSES IN STATISTICAL METHODS AND PLANNING OF INQUIRIES

Agri. Econ. 106.	Crop and Livestock Estimating
Agronomy 190.	Experimental Design
Commerce .	Industrial Statistics
Commerce 132.	(Econ. 132) Statistical Economics
Commerce 196.	(Econ. 196) Advanced Statistical Technique
Commerce 230.	(Econ. 230) Seminary in Statistical Research

Education 201.	Technique of Educational Research
Mathematics 135.	Introduction to Statistical Methods in the Natural Sciences
Mathematics 136.	Basic Principles of Process Control and Acceptance Inspection
Psychology 131.	Advanced Psychometric Methods
Psychology 231.	Design of Psychological Experiments
Sociology 182	Social Statistics: Correlation
Sociology 183.	Social Statistics: Sampling
Sociology 234.	Statistics in Population Research
Sociology 236.	Statistical Research in Sociology

COURSES IN PROBABILITY AND MATHEMATICAL STATISTICS

Mathematics 118.	Introduction to the Theory of Probability
Mathematics 137a.	Elementary Mathematical Statistics
Mathematics 137b.	Elementary Mathematical Statistics
Mathematics 208.	Probability and Analysis
Mathematics 231.	Foundations and General Methods of Probability

OTHER COURSES WITH STATISTICAL CONTENT

Agri. Econ. 155.	Prices of Agricultural Products
Agri. Econ. 255.	Seminary: Price Analysis
Education 129.	Tests and Measurements in Physical Education
Education 225.	Test Construction
Education 230.	Seminary: Measurement and Clinical Practice
Geography 126.	Cartography and Graphics
Physical Ed. 230.	Seminary in Measurement
Physical Ed. 262.	Research in Tests and Measurement
Psychology 135.	Psychology of Public Opinion
Psychology 143.	The Psychology of Individual Differences and the Measurement of Intelligence.
Psychology 230.	Seminary in Psychometric Methods
Sociology 133.	Quantitative Methods in Social Psychology
Sociology 235.	Research in Social Pathology

SELECTED COURSES IN MATHEMATICS

Mathematics 101a	Differential Calculus
Mathematics 101b.	Integral Calculus
Mathematics 106.	Advanced Analytical Geometry
Mathematics 114.	Advanced College Algebra
Mathematics 120.	Advanced Calculus
Mathematics 133.	Finite Differences and Interpolation
Mathematics 200.	Reading and Research (in Mathematical Statistics, for example)
Mathematics 220.	Theory of Analytic Functions
Mathematics 221.	Theory of Functions of a Real Variable

QUERIES

(*32*)

QUERY: In the elementary biostatistics books which I have consulted, the problem of skewed populations is generally dismissed with the warning that if the skewness surpasses a certain limit, then the procedures applicable to the normal distributions are no longer valid.

In attempting to quantify the so-called "courtship behaviour" of a certain tropical fish known as *Tilapia macrocephala,* I took

I had anticipated that the frequency of corded during a 15-minute interval would be normally distributed for the several spawnings. This, however, was not the case. With few exceptions the distributions resembled exponential curves, with the largest number of spawnings showing zero occurrence of the act for the interval in question, and in a few spawnings, the act was repeated a large number of times. An example follows:

Female "throat puffs" during the first prespawning interval

No. of throat puffs	0	1	2	3	4	5-9	10-19	15-19	20-24	25-29	30-34	35-39	40-44	45-49	*Total*
No. of spawnings	17	7	2	1	2	13	9	4	3	3	3	0	1	1	66

occurrence of the various behavioral acts recontinuous records of the occurrence of various aspects of this behavior from a variable time before spawning up to the spawning, and for one-half hour thereafter for a considerable number of spawnings. I then divided each record into 15-minute intervals starting with the spawning and working backwards. In this way I planned to determine the fluctuation of the various courtship acts.

The questions which arise are as follows:

1. Can the mean be used to describe the central tendency of such a distribution?

2. Can the range and/or standard deviation be used as measures of variability?

3. How can the significance of the differences between means be tested?

4. Can correlations be determined in the usual manner? (as for example, between the scores for males and females during a given

15-minute interval.) I notice that when a correlation table is constructed from these skewed distributions, one gets considerably more values in one corner of the table than in the diagonally opposite corner.

5. I have seen it suggested that the logarithm of the scores should be used to "normalize" markedly skew data. As it appears to me, this procedure tends to minimize the high scores, for which I see no justification as far as my data are concerned.

ANSWER: I sympathize with your feeling that the elementary texts have you on a limb in dealing with non-normal distributions. I don't know that the advanced texts would make you any happier. I think, though, that we can probably find some way of dealing with your distributions. I hesitate to make any flat general statements, because I feel that procedure must depend a good deal on the actual characteristics of the material and on what you are trying to do with it. So my remarks are to be regarded as tentative suggestions rather than as expert advice.

(1) You are, I take it, interested first in characterizing the frequency distributions, of which the one you give is a sample. Further, I take it that something on the order of 66 spawnings is your maximum number of observations. This probably means that measures of central tendency and scatter are all that are worth trying to compare. You would like to know what measures are likely to be most useful. You are rightly worried about the mean and standard deviation.

(2) There is clearly nothing to prevent your using means and standard deviations if you choose. The question, I think, is whether these are the best measures to use, in that there may be other measures which are subject to smaller sampling errors.

(3) One can, of course, use medians and percentile ranges. I think, however, that there is perhaps something better, namely, the use of some transformation which will at least approximately normalize the distributions. The transformations which seem most obvious are the use of $\log(1+x)$ or the use of $\sqrt{x}$.

(4) The quickest test of such normalizing transformations is to plot the cumulative distribution on probability paper. I have done this for your throat-puff distribution, and find that $\sqrt{x}$ gives, as I think, a very satisfactory normalization. If you find that it works equally on your other distributions, I think you can well take $\sqrt{x}$ as your variable, and determine means, standard deviations, and correlations of $\sqrt{x}$.

(5) The question you raise about the effect of transformation in "minimizing the high scores," and the justification of this, is one which can hardly be satisfactorily answered briefly. It would, I fear, take us quite into theory and philosophy. My view, if I had to compress it into a few lines, would be this: So long as I am concerned only with problems of location and scale, and so long as the relations I am finding are essentially empirical, I shall probably allow statistical convenience to govern the details of the statistical treatment. If and when I have a theoretical model for the phenomena under study, I may find that the model makes a particular statistical treatment appropriate.

C. P. Winsor

(33)

QUERY: Your statement in this *Bulletin*, Vol. 1, page 70, that the tabular probabilities must be doubled when the F table is used to test the homogeneity of variances makes me ask: Have you constructed a table of F values with 5% and 1% instead of 10% and 2% probabilities? If not, do you know of any place where such a table is available?

ANSWER: No, to both questions. It is easy to estimate the 5% and 1% points, with an accuracy sufficient for ordinary purposes, by linear interpolation. For greater precision, plot the tabular F values against the logarithms of their probabilities, then interpolate on a straight line. If you have the Fisher and Yates tables you can locate four points on a slightly curving line, then make your estimates still more nearly correct.

The ordinary tables of chi-square may be used for making the test. Two methods are available; one by E.J.C. Pittman (Biometrika, 31:200, 1939), the other by M. S. Bartlett (Supplement to the Journal of the Royal Statistical Society, 4:137, 1937).

G. W. Snedecor

(26)

JONES, D. F. (Connecticut Agricultural Experiment Station). **Heterosis and Synergesis.** (Paper read at St. Louis, March 27, 1946.)

In addition to the suppression of deleterious recessives and the accumulation of favorable dominants, there is the possibility of an interaction between alleles. Any interaction between non-allelic genes other than their dominance effect, probably has no bearing on heterosis since it should be manifest in the homozygous as well as the heterozygous condition. Recently a number of deviating lines have been found in inbred strains of maize whose expression is conditioned by a single gene. The alterations are both morphological and physiological and affect such characters as leaf width, height of stalk, internode size and shape, chlorophyll color and time of flowering. All but one of these variants, when crossed back to the normal line, gave significant increases in yield of grain above the more productive parent and matured earlier than the mean of the parents, in some cases earlier than the early parent.

The question immediately arises: is this heterosis within inbred lines due entirely to the single visible mutation or to other differences which may accompany this change resulting from multiple mutations or to delayed segregation. Mutations have appeared in controlled cultures in the form of visible morphological variations and in measurable physiological changes and it is equally possible that mutations also occur which have no noticeable effect in the homozygous condition but may have an appreciable effect in heterozygous combinations.

In the long inbred Leaming strains, started by East in 1905 and carried on by continuous self-fertilization, yield declined for 20 generations from above 80 bushels per acre at the start to 14, 22, and 24 for the five-year average of the three surviving strains up to the 20th generation. From the 20th to the 35th generation there was a further small decline to 11, 15, and 20 bushels and the long time trend is still downward. The yields obtained this past season which was exceptionally favorable for corn were 7, 16, and 19.

These inbred lines were separated in the ninth generation into two sublines each and maintained separately for seven additional generations. During this period, three lines showed no visible differences but when intercrossed within the line they all gave significant increase in some measurable character. Two of these lines were again separated after the 17th generation, further self-fertilized for six and eleven generations and again tested in intra-line crosses. No significant differences were obtained.

This evidence indicates that there is delayed segregation from an enforced heterozygous complex which has lasted for at least 15 generations. Since the inbreds in which the visible alterations appeared had not been continuously self-fertilized for this length of time, delayed segregation along with mutation seems to be the interpretation of the results obtained. An enforced heterozygous condition makes possible the accumulation of recessive mutations. In the usual method of maintaining commercial inbred lines minor segregations are usually eliminated before maturity, or are overlooked.

The deviating lines which give a hybrid vigor effect when crossed back to the normal line from which they originated appear to be single gene differences as far as they have been tested. Mono-factorial segregation is indicated by dwarf plant, pale top, and crooked stalk. Both the extracted homozygous normals and deviates have come out of the crosses slightly enlarged, as indication that other genes are involved. Further testing is necessary to establish the significance of these differences.

The normal lines in which the degenerate types appeared, in two cases tested, show no increases when crossed with the same normal line from which they have been separated for many generations. Therefore, the possibility of an accumulation of dominant genes from both parents in these two instances seems to be ruled out. If a normal gene *A* changes to recessive *a* and heterozygous *Aa* is greater than *AA* or *aa*, something other than the accumulation of dominant genes is indicated. Moreover, it makes little difference whether

one, or more than one, allele is involved since the change or changes are all in one direction if the normal line remains the same. If there is delayed segregation then heterozygous *AaBb* resolves into homozygous *AABB, AAbb, aaBB, aabb*, to take the simplest example. Recombination, if there is incomplete dominance, should make possible sublines both above and below the level of vigor of the parental line. Crossing any sublines with the original heterozygous parental line would not give an increase over the better parent on the basis of dominance but crossing the sublines would restore the original level. Actually there is an increase when the mutant lines are crossed with the original parents. This is evidence for an interaction between alleles. Final proof must come from the extracted recessives when again crossed with the normal line from which they deviated. These tests are now being made.

(27)

COLE, LaMONT C. (U. S. Public Health Service). **The Non-Random Distribution of Animal Populations — An Urgent Problem for Further Investigations.**

Assuming that if animals were randomly scattered in space their distribution would correspond to the Poisson exponential, it is found that nearly all natural populations are contagiously distributed, i.e., occupy less space but form larger aggregations than in a random distribution. This may result from sexual attraction, social instincts, sample heterogeneity, inappropriate sample size, or common origin of the individuals in each group. In populations, both of certain free-living arthropods and of rodent-ectoparasites, it is easily demonstrated, using quantities such as the percentage of fertile samples or the Charlier coefficient of disturbance, that the extent of the departure from randomness may vary systematically with time and may be correlated with meteorologic and other environmental conditions.

Contagious distribution hampers population analysis by restricting the range of applicable statistical procedures, by leading to large variances which reduce the sensitivity of significance tests, and by introducing a bias into census estimates made by sampling.

In the material investigated by the writer, the theories of Polya and of Neyman have not been applicable — their forms generally failing to fit the empirical distributions even when the parameters are so adjusted as to give perfect results on the first few frequency classes. The most hopeful approach found by the writer derives from the seemingly rational hypothesis that a contagious distribution may be interpreted as an intermingling of random distributions of various-sized groups of organisms. From this hypothesis good fits are obtained on a wide variety of contagiously distributed populations. If future experience shows this hypothesis to be tenable, a variety of new procedures for population analysis may become available.

(28)

COMIN, DONALD and WESLEY P. JUDKINS. (Ohio Agricultural Experiment Station). **The Suitability for Freezing of Varieties of Sweet and Sour Cherries.**

A statistical study was made of fourteen sweet and seven sour cherry varieties as to their suitability for freezing. A testing jury of five Experiment Station research men and five women of the Home Economics Department separately sampled the cherries which had been frozen with and without sugar solution for one year. Each judge rated each variety under any one set of conditions on a numerical basis. The ratings were then ranked from one to fourteen (in the case of fourteen varieties) and the scores subjected to a normalizing transformation similar to the technique suggested by C. I. Bliss in Storrs, Connecticut, Agricultural Experiment Station Bulletin 251. Each test was then examined by the analysis of variance to determine the judges' preference for varieties, type of pack and sex differences exhibited by the judges.

This analysis showed the judges differentiated between varieties, the men more than the women, to the extent that both sexes showed a significant preference for some varieties over others. There was a tendency for all judges to group the varieties into classes, the members of which were quite similar in their suitability for freezing. For example, all yellow sweet cherry varieties were found to

bear scores close to each other and be definitely inferior to the red varieties. Thus it was possible to point out the superiority of cherries frozen in 50 per cent sugar solution to those frozen without sugar, and to rate all varieties in a classification of superior, excellent, fair (or acceptable) and poor or unacceptable.

(29)

SNEDECOR, GEORGE W. and E. S. HABER. (Iowa State College). **Statistical Methods for an Incomplete Experiment on a Perennial Crop.**

In 1939 the scarcity of labor terminated an experiment on asparagus which was planned to run for 20 years or longer. Data from 1929 to 1938 were available on three harvesting dates in six randomized blocks, the three harvests ending on June 1, June 15, and July 1.

Linear and parabolic trends indicated: (i) that cutting to June 1 failed to utilize the potential yield; (ii) that cutting to July 1 depleted the reserves of the plants so that they were either dead or uneconomical by 1938; and (iii) that cutting to June 15 was close to optimum.

By fitting a parabola to the accumulated yields from the three treatments each year, it was found that the date of maximum yield receded from July 18 in 1932 to June 19 in 1938. The curve of recession was hyperbolic, and the curve was used to forecast optimum dates. These approached June 17 in 1955. At the 20th harvest in 1948, the optimum date is estimated as June 17.5 with the 50% confidence interval 17.2-18.0.

(30)

FEDERER, W. T., and G. F. SPRAGUE. (Iowa State College). **The evaluation of error, tester x line and line components of variance in corn top-cross experiments. J. Amer. Soc. Agron.**

The utilization of hybrid vigor in either plant or animal breeding requires the evaluation of inbred lines in hybrid combinations. The top-cross method (line by tester) has been used extensively for the preliminary evaluation of combining ability. The question arises as to the adequacy of a single tester.

In reference to this question statistical investigations were conducted on 11 experiments which involved lines of corn, testers and replicates. The error, the tester x line and the line components of variance were not homogenous. Unbiased estimates of the ratios of the tester x line to the error components of variance and of the line to the error component were obtained for the eleven experiments. The average tester x line component was about 23 percent and the line component was about 57 percent of the error variance component.

Yates' formula for the average genetic advance due to the selection of the apparently best line instead of a random line was extended to include the case in which the experimental error for a line mean contained two components. Furthermore, it was found that it may be more profitable on the average for the corn breeder to use more testers. The highest average genetic advance due to the selection of the apparently best line or lines instead of a random one or ones was found to depend upon the number of lines, the number of testers and the number of replicates.

(31)

RIGNEY, J. A. (North Carolina State College). **Some Statistical Problems Confronting Horticultural Investigators. (Paper read before joint session of Biometrics Section and American Society of Horticultural Science at AAAS meetings, St. Louis, 1946).**

Three aspects of planning experiments in biological research were considered. Choice of treatments affects the efficiency of a test. Where considerable information has already been obtained on the general nature of interactions to be encountered, a complete factorial may not be the most efficient combination of treatments to use. Under some circumstances, the most useful information may be obtained from testing various levels of each element under conditions where it alone is limiting.

A second problem is that of sampling. Soil sampling procedures are in need of statistical evaluation. Sampling of plant tissues for chemical analysis needs study also. Statistical techniques for the former problem are inadequate; for the latter type of problem, they are available and merely need extensive application.

The third problem discussed is the choice of suitable designs. The general usefulness

of randomized block design was stressed. For field experiments, the use of incomplete block designs which cannot be used alternatively as a randomized complete block is questioned. Data from corn variety tests in the Southeast indicate an average relative efficiency of about 150 for various kinds and sizes of experiments. Thirty-five 4 x 4 lattices averaged 142 which is much greater than that reported from other sections. Many conditions in the Southeast are peculiarly suited to the use of lattice squares. Some of these are described.

(32)

WADE, B. L., FRANCES L. HAYDEN and P. H. HEINZE. (U. S. D. A., Regional Vegetable Breeding Laboratory). **The Lattice Square Applied to Vitamin Content of Snap Beans.**

Data on ascorbic acid from a 7 x 7 lattice square for spring and fall crops gave some increase in precision when the later Yates method of analysis was used but not so great as that for yield. The gain in precision for yield of lattice square designs compared to randomized blocks was 65% for spring and 22% for fall while the gain in precision for ascorbic acid content did not in any case exceed 25%. For thiamine and riboflavin, laboratory capacity did not permit the use of the lattice square but the coefficients of variability encountered for these two vitamins indicated that the gain in precision by use of the lattice square might be similar to that for ascorbic acid. The later Yates method of adjusting mean yields, in which the information contained in the incomplete blocks is utilized, was satisfactory when aplied to very variable fall stands and yields, in conjunction with covariance techniques whereby yield was adjusted for stand. Satisfactory adjustments of mean yields for the fall crop could not be made if the data were analyzed as randomized blocks or by the earlier Yates method in which no separate corrections were made for row and column differences.

(33)

GOWEN, JOHN W. (Iowa State College). **Hybrid Vigor in Drosophila. (Paper presented at the joint session of the American Statistical Association and the Genetics Society of America at St. Louis).**

The hybrids between inbred races of Drosophila have shown egg yields between equality with that of the parents and three times that of the parents. The heterotic effect extends to the characters duration of life of both males and females. In egg production the heterotic effect is expressed in all the F_1's causing all of them to become good producers rather than simply making a few of them superlative performers. Hybrid vigor in egg production is about equally expressed by records early in the life of the female as by her full life production. The production curve is such that the hybrid increased yield over the inbred parents could be interpreted as due to reserves of metabolic products accumulated before the egg production actually commenced. These products are expended over the egg laying period at about the same rate for the hybrid as for its inbred parents. The difference comes in the fact that the hybrids start production at higher average levels.

The separation of genic from cytoplasmic effects indicates that hybrid vigor is due to gene differences and not the union of different cytoplasms. A genom from an inbred line may show general combining ability with many different genoms of other lines or specific combining ability with certain genoms. Experimental analysis of the relative influence of these two types of effects shows them to be about equally important to hybrid vigor. The genom contribution may be further subdivided into the contribution of its individual chromosomes. A study of these effects shows that the relation between heterozygosis and hybrid vigor is linear. No significant interaction effects between chromosomes exist. Significant heterotic effects were found for each chromosome. The heterotic effect was proportional to the active length of the chromosome as measured by per cent of visible loci, length in cross-over units or by bands in the salivary gland chromosomes. Both plus and minus genes are located in all chromosomes. A minimum of two or three genes per chromosome appear necessary to hybrid vigor. Analysis through back-crossing and selection indicates that genes with heterotic effects in egg production are dominant or nearly so with additive action between separate pairs.

University College, London, has statistical activities in the Department of Biometry and in the Galton Laboratory. J. B. S. HALDANE is head of the Department of Biometry and editor of the Journal of Genetics. His research interests include the genetics of man, culex and drosophila with statistical methods. The other members of the staff with their research fields follow: MYRTLE J. BYRD, genetics of drosophila subobscura; H. GRUNEBERG, genetics of the mouse; URSULA PHILIP, cytology of drosophila species; HELEN SPURWAY (now Mrs. Haldane), genetics of drosophila subobscura and amphibie and A. C. ELIZABETH SULEY, secretary and research assistant. L. S. PENROSE is a Galton professor and editor of Annals of Eugenics; also, he is doing research in human genetics, congenital malformations, biochemical abnormalities, and autosomal linkage. His staff includes HANS KALMUS,genetics of drosophila. carbon dioxide sensibility inheritance, human genetics, tone deafness, carcinoma of the breast, and toxaenmia of pregnancy; M. N. KARN, assistant editor of Annals of Eugenics; S. B. HOLT, inheritance of dermatoglyphs; and B. E. SIMPSON, secretary to the Laboratory . . . Some other English scientists interested in statistics as applied to biological research are A. G. POLLARD, Imperial College of Science, S. Kensington, London; O. V. S. HEATH, Research Institute of Plant Physiology, Imperial College, London, K. MATHER, John Innes Horticultural Institution, London; M. L. SEMPLE, Jealotts Hill Research Station, Bracknell, Berks; J. O. IRWIN, London School of Hygiene; and JOSEPH EDWARDS, Head of the Department of Dairy Husbandry, Milk Marketing Board, Surrey, England . . . In Scotland we hear about A. D. BUCHANAN-SMITH and H. P. DONALD who are with the Institute of Animal Genetics, University of Edinburgh. At the same University is J. H. GADDUM in Pharmacology from whom we have heard . . . H. C. FORSTER, Chairman, War Agricultural Committee, Department of Agriculture, Victoria, Australia, writes, "Things have been pretty mixed here over the past six years, and I haven't had the chance to do much more than just glance at a bit of statistics. During the war most research in the State Department came to a standstill, and I was drafted to set up and look after the War Agricultural Committees in Victoria. The direction of over 250 such committees with all their worries of manpower, machinery, and materials for a maximum food production effort has kept me more than busy." . . . R. D. CROLL, who worked for four years with Mr. Forster on agricultural experiment design in the Victoria Department of Agriculture, is now with the Drug Houses of Australia, Ltd., Melbourne. He is in charge of the Research Division of this firm which manufactures and merchandizes industrial chemicals and pharmaceutical preparations, many of which are derived from biological sources. He states, "We exercise a good deal of quality control in the statistical sense and carry out large-scale field tests on botanical drug plants. We also derive some of our raw materials from milk and animal tissues. Another interest is the biological testing of drugs and the testing for efficiency of insecticides and weedicides." . . . HELEN NEWTON TURNER is at the McMaster Animal Health Laboratory, Glebe, Sydney, which is a Division of the Council for Scientific and Industrial Research. She writes, "I've been trying to persuade our University to start a Statistics Department." They are having an influx of students as in our American Universities. As she states, "More than double the pre-war figure." . . . D. L. SERVENTY is with the Fisheries Department at Perth, Western Australia, another section of the Council . . . No longer is it F. E. Allen, but MRS. P. J. CALVERT. She is engaged now as a part-time research officer in Demography at the Commonwealth Bureau of Census in Statistics, as lecturer in Statistics and consultant at the Commonwealth Forestry Bureau . . . H. SCHWERDTFEGER, The University of Adelaide, Department of Mathematics, sent a notice of the, "Colloquium on the theory of the Gamma-function and its applications," being conducted at the University. The lectures were given to meet the interests of students of Mathematical Statistics . . .

Another Professor of Mathematics, M. H. BELZ, at the University of Melbourne, Victoria, writes that he is interested in the development of mathematical statistics. MILDRED BARNARD PRENTICE is on his staff . . . JOHN H. HARDING, Nedlands, Western Australia, writes "We have been employing statistical methods in Forestry and Agricultural Research here for some considerable time." . . . We would like a note from W. W. BRYAN, Queensland Agricultural College. Are you using the incomplete block or lattice designs in your plant breeding program? . . . We would also like to hear from P. P. McGOVERN, Queensland Department of Agriculture and Stock; F. C. McCLEERY, Department of Agriculture, New South Wales; and E. WILLIAMS, C.S.I.R., Melbourne . . . J. T. CAMPBELL, Victoria University College, Wellington, New Zealand, is a lecturer in Mathematics. He states, "Since working under A. AITKEN, Edinburgh University, Scotland, statistical mathematics is my special interest." Mr. Campbell feels that there is a tremendous field ready in New Zealand for the development and use of statistical methods, especially in agriculture . . . D. B. DUNCAN, on a traveling scholarship from the University of Sydney, Australia, has just completed a year's work at Ames, under G.W. SNEDECOR, W.G. COCHRAN and A.M. MOOD. At present he is attending the special Summer Session at the Institute of Statistics, North Carolina, and is taking the courses with R. A. FISHER, GERTRUDE COX, and J. WOLFOWITZ. He plans to return to Ames in the late summer and will leave for England early in the new year to study at Rothamsted and Cambridge.

A series of Genetics Seminars was organized during the statistical summer session in order to hear R. A. FISHER'S views on current genetic topics and to give him the opportunity of meeting local plant breeders. One of the major topics considered was the measure of dominance in corn. F.H. HULL, Agronomist at Florida Agricultural Experiment Station, presented his new techniques for measuring of dominance and "over-dominance." P. H. HARVEY, in charge of corn breeding, North Carolina State College, prepared the group for Mr. Hull's discussions by outlining the usual breeding techniques currently used in corn. THOMAS KERR, in charge of cotton breeding, North Carolina Agricultural Experiment Station, reviewed the available information on the origin of species of cotton and indicated some of the problems facing cotton breeders. KENNETH DODDS, Cytogeneticist, Imperial College of Tropical Agriculture, Trinidad, gave a review of banana breeding and an interesting example of the evolution of dominance under controlled selection. R. A. Fisher, Cambridge University, England, gave a talk on the types of gamete formation possible in polyploids. B. W. SMITH, Cytogeneticist in forage crops, North Carolina State College, reviewed the use of trisomics in genetic investigation and C. I. BLISS, Yale University, presented an example of the use of trisomics in the investigation of vitamin B in the production of corn. The seminar was attended by approximately 50 local and visiting geneticists. Some of the visitors were H. L. BUSH, Statistician, The Greatwestern Sugar Company, Colorado; J. F. CROW, Biostatistics, Dartmouth College; H. O. HETZER, Regional Swine Breeding Laboratory, Washington, D. C., J. H. WEATHERSPOON, Plant Breeder, Pioneer Hibred Corn Company.

Material for the BULLETIN should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C., material for Queries should go to "Queries," Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.